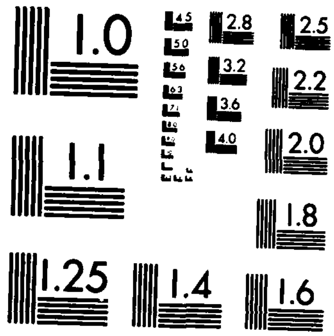
AD-A141 389 A METHOD FOR CONSTRUCTING A UNIMODAL INFERENTIAL OR PRIOR DISTRIBUTION(U) TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES P BROCKETT ET AL. JAN 84 CCS-RR-473 1/1

UNCLASSIFIED N00014-81-C-0236 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Research Report CCS 473

A METHOD FOR CONSTRUCTING A UNIMODAL INFERENTIAL OR PRIOR DISTRIBUTION

by

P. Brockett
A. Charnes
K. Paick

CENTER FOR CYBERNETIC STUDIES
The University of Texas
Austin, Texas 78712

Research Report CCS 473

A METHOD FOR CONSTRUCTING A UNIMODAL INFERENTIAL OR PRIOR DISTRIBUTION

by

P. Brockett
A. Charnes
K. Paick

January 1984

Acknowledgement: We wish to thank D.V. Lindley for encouraging comments on the ideas presented in this paper.

This research was partly supported by ONR Contracts N00014-81-C-0236 and N00014-82-K-0295 with the Center for Cybernetic Studies, The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Graduate School of Business 4.138
The University of Texas at Austin
Austin, Texas 78712
(512) 471-1821

Abstract

In this paper we show how to take personally assessed information and use it to develop a continuous unimodal prior density function, perhaps for subsequent Bayesian analysis. The method is completely nonparametric and uses only the furnished information and no other. The technique is easily computerized, and yields a closed analytical formula for the prior. The resulting distribution may be considered to be an inferential distribution.

Keywords: Prior distribution assessment, unimodality, information theory, maximum entropy

1. Introduction

A basic rule in statistical analysis is to use all the information which you have, but avoid using any information which you don't have. The desire to use the prior belief or experiences of the scientist as information capable of being input into the statistical analysis has led to the branch of Bayesian statistics, and the same goal has also led to information-theoretic statistical analysis. One crucial problem which must be addressed in applying any Bayesian methods in real situations in business, medicine, and other fields is how to take the information supplied by the client (or the scientist himself) and obtain an inferential (or prior) distribution for the stochastic phenomenon under study. In particular, an inferential distribution must be found in order to update expectations (and find Bayes estimators), keeping in mind that you should avoid using information which you don't have. In real applications you should not necessarily assume a parametric prior (such as a normal prior) and just proceed to estimate parameters unless the distributional model has been given as part of the information. To paraphrase Albert Einstein, the model should be as simple as possible, but no simpler.

In this paper we address the topic of inferential density assessment when we know the prior density is unimodal. We quantify the amount of information in a statistical density by using the information-theoretic techniques, and we show explicitly how to "use all the information available (including unimodality) and no other". Our technique involves transforming the problem from the original unimodal stochastic variable to an auxiliary variable. We then estimate the density for the auxiliary variable using minimum discrimination information subject to the constraints obtained about the original variable. The detailed formulae are given in the next several sections. The result is easily computerized so that the user need only input a few basic characteristics, and the computer then outputs a graphical and also analytical representation of the desired density.

2. Information Theoretic Density Estimation

In information theoretic notation the expected amount of information in an observation X for distinguishing between two density functions f and g is denoted by $I(f|g)$. Mathematically this expected information is quantified by the expected value of the difference in log-odds ratio or Kullback-Leibler number, viz,

$$I(f|g) = \int f(x)\, ln\left[\frac{f(x)}{g(x)}\right] \lambda(dx) \tag{2.1}$$

where λ is some dominating measure for f and g. We shall call $I(f|g)$ the informational divergence between f and g. If $g(x) = 1$ for all x, then $I(f|g)$ represents the informational divergence between the postulated density f and the completely uninformative density g. In this case $I(f|g)$ is precisely minus the entropy of f, and is a measure of the uncertainty of the density f. In practice, $\lambda(dx)$ is selected as Lebesgue measure in the (absolutely) continuous case, and as counting measure in the discrete case.

If we are given certain generalized moment constraints which the density f must satisfy, such as

$$\begin{aligned} 1 = b_0 &= \int f(x)\lambda(dx) \\ b_1 &= \int h_1(x) f(x) \lambda(dx) \\ &\;\;\vdots \\ b_k &= \int h_k(x) f(x) \lambda(dx) \end{aligned} \tag{2.2}$$

then the minimum distrimination information (MDI) estimate of the density g subject to the constraints (2.2) is defined to be the minimum of (2.1) over all f subject to (2.2). If $g(x) = 1$, then this density estimate is called the maximum entropy density (M.E.) subject to (2.2). This density is the least informative distribution possible subject only to the constraints (2.2). If we are to use only that information given in (2.2), and no other information, then the M.E. density is implied. This constrained M.E. density estimation may be construed as a useful extension of Laplace's famous "principle of insufficient reason" which postulates a uniform distribution in the situation in which no knowledge is available. Here, when information of the form

(2.2) is available, we select that distribution which is as close to uniform as possible subject to the given constraints (2.2). Of course other "goal densities" g may be more appropriate in other situations.

The explicit calculation of the MDI density subject to (2.2) is easily carried out using Lagrange Multipliers. Introducing a multiplier a_i for constraint i, we wish to maximize $-I(f|g)$. We have

$$-I(f|g) = \int f(x)\, ln \left[\frac{g(x)}{f(x)}\right] \lambda(dx) - \sum_{i=0}^{k} a_i b_i$$

$$= \int f(x) \left\{ ln\left[\frac{g(x)}{f(x)}\right] - \sum_{i=0}^{k} a_i h_i(x)\right\} \lambda(dx)$$

$$= \int f(x)\, ln \left[\frac{exp\left\{-\sum_{i=0}^{k} a_i h_i(x)\right\}}{f(x)} g(x)\right] \lambda(dx)$$

$$\le \int f(x) \left[\frac{exp\left\{-\sum_{i=0}^{k} a_i h_i(x)\right\}}{f(x)} g(x) - 1\right] \lambda(dx)$$

where $h_0(x) = 1$. The inequality follows since $\ln x \le x - 1$ with equality only when $x = 1$. Thus the above inequality becomes an equality when

$$f(x) = exp\left\{-\sum_{i=0}^{k} a_i h_i(x)\right\} g(x).$$

Summarizing, the MDI density subject to the constraints (2.2) is precisely

$$f(x) = exp\left\{-\sum_{i=0}^{k} a_i h_i(x)\right\} g(x) \tag{2.3}$$

where $h_0(x) = 1$, and the constants a_i, $i = 0, 1, \ldots, k$ are found by solving the moment constraints (2.2) simultaneously.

An easier method for determining the actual numerical values for a_i can be derived from the results in Brockett, Charnes and Cooper (1980), or in Charnes, Cooper and Seiford (1978). There it is shown that the problem of minimizing (2.1)

subject to (2.2) is a constrained strictly convex programming problem with an unconstrained dual convex program involving only exponential and linear terms. Moreover, the desired a_i, $i = 0, 1, \ldots, k$ are precisely the dual variables and may easily be obtained on a computer by any of a number of nonlinear programming codes (e.g., the SUMT method code of Garth McCormick and Charles Mylander). This duality relationship is exploited in the final section where numerical examples are given. Thus the entire process of obtaining the MDI density is easily computerized.

3. Exploiting Knowledge About Unimodality

Most people when forecasting or estimating a prior distribution will have a unimodal density in mind. Consequently the M.E. density obtained in section 2 will be rejected by most decision makers, since, except for certain serendipitous situations concerning the relationships among the functions $h_i(x)$ and the b_j's, the M.E. density cannot be guaranteed to be unimodal. A bi- or tri-modal density is hard to justify to the decision maker in many (but not all) situations. Accordingly, in this section, we show how to incorporate the knowledge that the prior density for the decision maker is unimodal. This technique is of independent interest in improving many procedures to incorporate unimodality. We shall show how to obtain an inferential (or prior) distribution which is as uninformative as possible subject to being unimodal and satisfying the constraints (2.2)

As a concrete example let us suppose we have elicited from the decision maker the following information concerning an unknown prior variable θ: (these may be as the result of sales forecasts for example).

1) The prior density for θ is unimodal with the most likely value θ_0.
2) The range of possibilities for θ is a to b.
3) The decision maker will give even odds that θ is between two numbers a_1 and a_2. (This will give a measure of dispersion for the desired density. Prescribing the 25^{th} and 75^{th} percentiles is another usual vehicle for obtaining this sort of measure.)
4) The decision maker assesses the chance of θ falling short of θ_0 as p. (This will give a measure of skewness for the prior density).

These constraints translate into:

i) θ is unimodal with mode θ_0

and

ii)

$$1 = b_0 = \int_a^b h_0(\theta) f(\theta)\, d\theta$$

(3.1)
$$.5 = b_1 = \int h_1(\theta) f(\theta)\, d\theta$$

$$p = b_2 = \int h_2(\theta) f(\theta)\, d\theta$$

where

$$h_0(\theta) = 1, \quad h_1(\theta) = \begin{cases} 1 \; if\, \theta \in [a_1, a_2] \\ 0 \; if\, \theta \notin [a_1, a_2] \end{cases}$$

and

$$h_2(\theta) = \begin{cases} 1 \; if\, \theta \le \theta_0 \\ 0 \; if\, \theta > \theta_0 \end{cases}.$$

Of course other possible constraints (such as other assessed percentiles) may be added if the statistician desires.

The problem addressed in this section is how the Bayesian statistician (or more generally, the scientist desiring an inferential distribution) can use the unimodality information in a constructive manner.

If θ is a unimodal variable with mode θ_0, then the density satisfies $f'(\theta) \ge 0$ for $\theta < \theta_0$ and $f'(\theta) \le 0$ for $\theta > \theta_0$. Accordingly $-(\theta-\theta_0)f'(\theta) \ge 0$ for all θ, and so $-(\theta-\theta_0)f'(\theta)$ is proportional to the density of some random variable, X. This is equivalent to the decomposition

(3.2)
$$\theta - \theta_0 = U \cdot X$$

where U is uniform over [0,1] and independent of the random variable X, which, of course, is just L. Shepp's reformulation of Khinchin's famous characterization of unimodal random variables (cf. Feller (1971) page 158).

Since knowledge that θ is θ_0 - unimodal is completely equivalent to the existence of the decomposition (3.2) we know such a random variable X must exist. However we are ignorant about its precise form except for the constraints upon X which are implied by the known constraints (2.2) on θ. By using the decomposition (3.2), together with conditional expectation given X we obtain

(3.3)
$$E[\,h(\theta)\,] = E[\,h^*(X)\,]$$

where

$$h^*(x) = E[h(UX + \theta_0) | X = x] = \frac{1}{x}\int_0^x h(t + \theta_0)\,dt.$$

The relationship (3.3) allows us to transform the moment constraints (2.2) on θ into moment constraints of the form

$$b_i = \int h_i^*(x) f_X(x)\,dx, \quad i = 0, 1, 2, \ldots, k$$

upon X. In this regard the transformation (3.2) is in the spirit of Kemperman (1971).

As an illustration, the transformed constraints (3.1) become

$$1 = b_0 = \int h_0^*(x) f_X(x)\,dx$$

$$5 = b_1 = \int h_1^*(x) f_X(x)\lambda(dx) \qquad (3.4)$$

$$p = b_2 = \int h_2^*(x) f_X(x)\lambda(dx)$$

where

$$h_0^*(x) = \frac{1}{x}\int_0^x dt = 1,$$

$$h_1^*(x) = \begin{cases} 0 & \text{if } x < a_1 - \theta_0 \\ 1 & \text{if } a_1 - \theta_0 \le x \le a_2 - \theta_0 \\ \dfrac{a_2 - \theta_0}{x} & \text{if } x \ge a_2 - \theta_0 \end{cases}$$

and

$$h_2^*(x) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } x > 0 \end{cases}$$

The problem of constructing an inferential (or prior) distribution for θ subject to the constraints (3.1) and unimodality has now been transformed via (3.2) and (3.3) into the problem of constructing a density function for X subject to the constraints (3.4). By unimodality of θ we know X exists, however we have no information concerning X other than the fact that it satisfies (3.4). Accordingly, we may use the

extension of Laplace's principle of insufficient information to postulate a maximum entropy density for X. We obtain (from (2.3))

$$f_X(x) = exp\left\{ - \sum_{i=0}^{k} a_i h_i^*(x) \right\} \quad for\, x \in [a - \theta_0, b - \theta_0], \tag{3.5}$$

and again using the relation (3.2) the original density for θ becomes

$$f(\theta) = \begin{cases} \int_{-\infty}^{\theta - \theta_0} f_X(x) \frac{dx}{|x|} & if\, \theta < \theta_0 \\ \int_{\theta - \theta_0}^{\infty} f_X(x) \frac{dx}{x} & if\, \theta > \theta_0 \end{cases} \tag{3.6}$$

The constants $\{a_i\}$ needed to determine (3.5) and hence (3.6) are found using the unconstrained dual formulation discussed previously which is explicitly stated in the final numerical example section.

The computer program to implement this analysis should also plot the obtained inferential or prior density using the construction (3.6). One might use this graph for consultation with the decision maker. If more information is available or needs to be supplied for decision making purposes, additional constraints are added to (3.1), transformed into new constraints on X via (3.3) and added to the constraint set (3.4). Such supplementation can be continued until the decision maker when presented with the graphical density representation is satisfied with the inferential density obtained.

4. Numerical Illustrations

In this section we shall exhibit the numerical results of implementing the previous procedure. From the duality theory given in Charnes, Cooper and Seiford (1978), the dual to the primal problem of minimizing (2.1) subject to the constraints (2.2) is the unconstrained convex programming problem

$$\max_{a} \sum_{i=0}^{k} a_i b_i - \int g(x)\, exp\{- \sum_{i=0}^{k} a_i h_i(x)\} \lambda(dx). \tag{4.1}$$

In the unimodal estimation problem considered in the previous section, one estimates the parameters $\{a_i\}$ in the density $f_X(x)$ by solving (4.1) with h_i^* replacing h_i, and $\lambda(dx) = dx$.

One further point should be made here. If one wishes to impose a continuity constraint upon the density f_θ at the mode θ_0, then, from equation (3.6), such a constraint on θ translates into a moment constraint upon the auxiliary variable X of

the form

$$0 = \int h^*_{k+1}(x) f_X(x)\,dx \tag{4.2}$$

where $h^*_{k+1}(x) = 1/x$ (there is no corresponding h_{k+1} constraint on θ). Similarly, smoothness constraints upon the density f_θ near the mode θ_0 can be accomplished by choosing "goal densities" $g(x)$ in the estimation of the auxiliary density f_X which are sufficiently smooth as $x \to \theta_0$. We shall illustrate this with three goal densities. Up to the appropriate normalizing constant these are $g_1(x) = 1$ (corresponding to maximum entropy estimation for f_X)

$$g_2(x) = \begin{cases} exp\{-(|x|-\delta)^2 - (\frac{1}{|x|} - \frac{1}{\delta})^2\} & for\ |x| \le \delta \\ 1 & for\ |x| \ge \delta \end{cases}$$

and

$$g_3(x) = exp\{-x^2/2\sigma^2\}/\sqrt{2\pi}\sigma$$

The goal density g_2 behaves like the constant 1 outside $|x| \le \delta$, and dips smoothly to zero as $|x| \to 0$. This goal density approximates the maximum entropy procedure given by g_1, but constrains the resulting prior density f_θ to be smooth at the mode. The goal density $g_3(x)$ corresponds to the f_X density which would result from f_θ being normally distributed, and hence this goal density gives the "close to normality subject to constraints" result for the estimated prior density f_θ.[1]

Figure 1 shows the resulting prior distributions obtained using each of these goal densities and using only the following client furnished information concerning θ:

1. θ is unimodal with possible values between 0 and 10

2. the most likely value for θ is 3

3. there are even odds that the value of θ lies between 1 and 5

[1] The continuity and smoothness at the mode is not guaranteed for asymmetric density using the goal density $g_3(x)$, due to the symmetry of the normal. To impose the smoothness and the "close to normality subject to constraints" interpretation, the product of $g_2(x)$ and $g_3(x)$ can be used as a goal density. Figure 1 shows this case.

4. there is a 30% chance that θ will fall short of the most likely value of 3.

(INSERT FIGURE 1 HERE)

It can be seen that the parameter δ in the goal g_2 serves as a smoothing parameter.

For a second illustration, assume we have the following information:

1. θ is unimodal with possible values between 0 and 10

2. The most likely value for θ is 5

3. there are even odds that the value of θ is between 4 and 6

4. the distribution of θ is symmetric about 5

(INSERT FIGURE 2 HERE)

Figure 2 shows the results of the calculations in this second situation for each of the goal densities.

As a final note, we should remark that the technique described in this paper can also be extended into a new technique for non-parametric unimodal density estimation using actual data and prior information. We develop this non-parametric unimodal density estimation technique in Brockett, Charnes and Paick (1983).

References

Brockett, P.L., A. Charnes and W.W. Cooper, (1980) "MDI Estimation via Unconstrained Convex Programming", Commun. Stat.-Simula. and Computa, B9(3), 223-234.

Brockett, P.L., A. Charnes and K. Paick, (October 1983) "Information-Theoretic Non-Parametric Unimodal Density Estimation", Research Report, CCS report 467 Center for Cybernetic Studies, University of Texas, Austin, Texas (in process).

Charnes, A., W.W. Cooper and L. Seiford, (1978) "Extremal Principles and Optimization Dualities for Khinchine-Kullback-Leibler Estimation", Math. Operationsforsch. Statist. Ser. Optimization 9, no. 1, 21-29.

Feller, W. (1971) An Introduction to Probability Theory and Its Application 2. New York; Wiley.

Kemperman, J.H. (1971) "Moment Problems with Convexity Conditions I", in Optimizing Methods in Statistics, Academic Press. p 115-178.

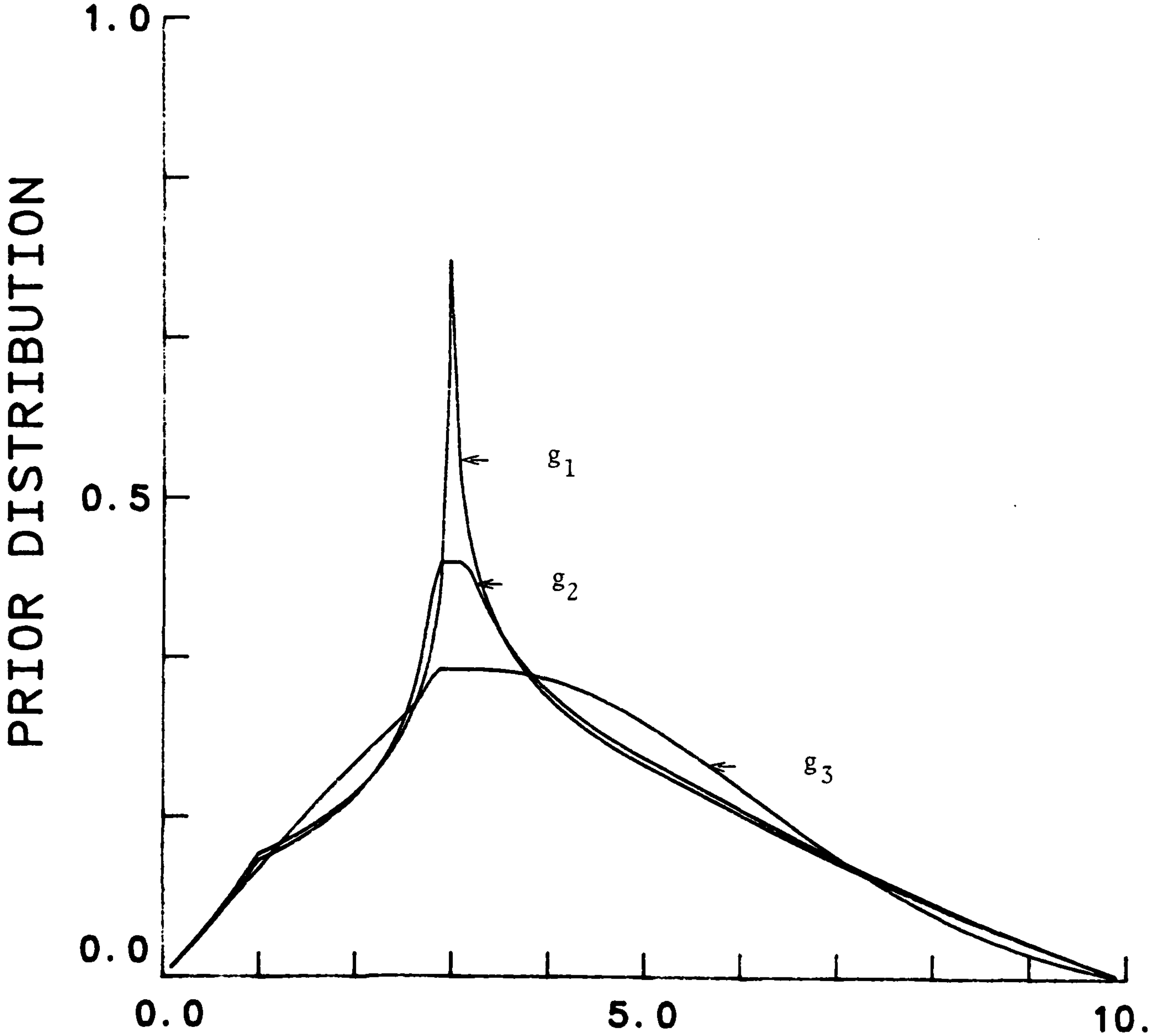


FIGURE 1

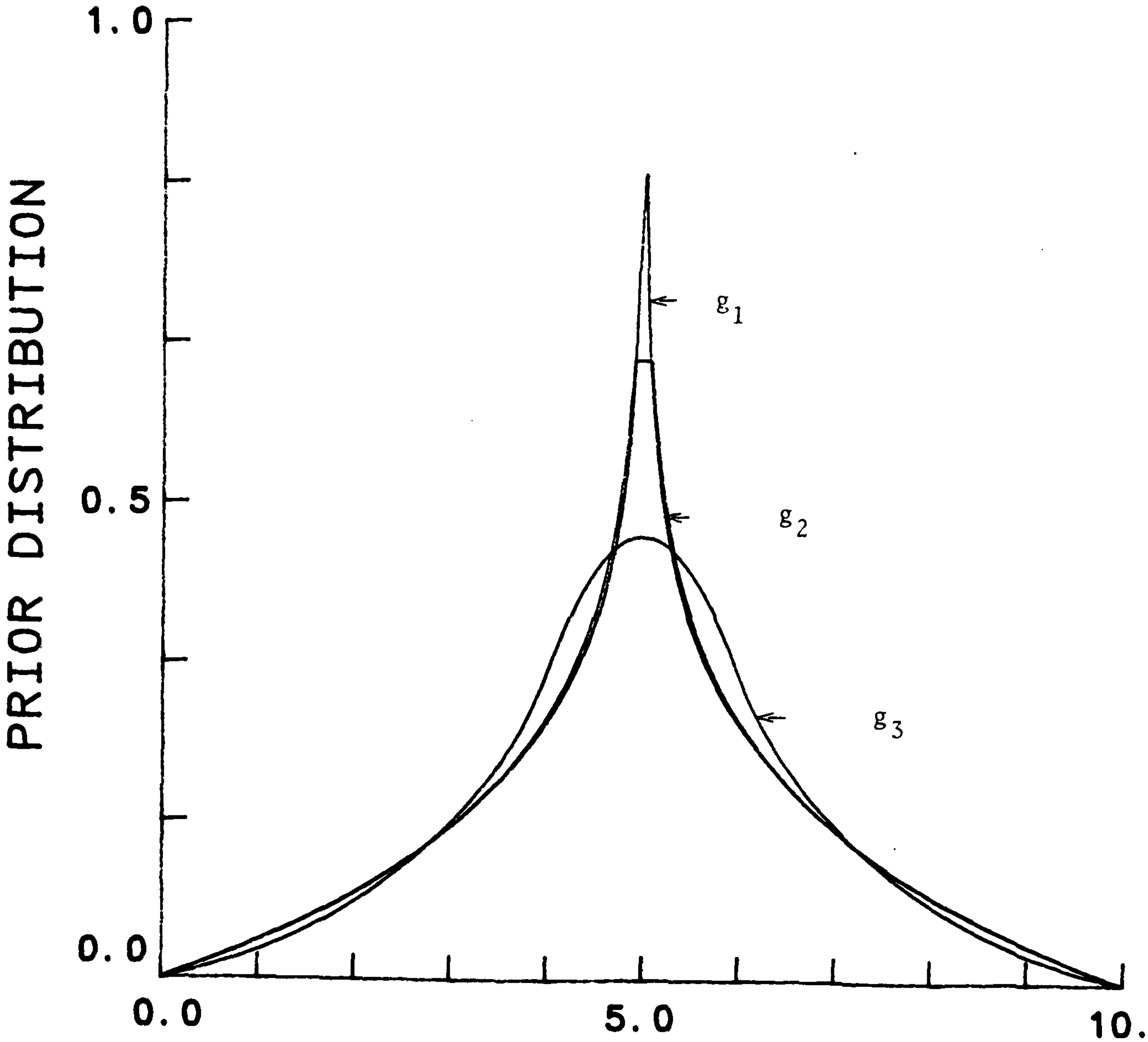


FIGURE 2

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CCS 473	2. GOVT ACCESSION NO. AD-A141387	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Method for Constructing a Unimodal Inferential or Prior Distribution		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) P. Brockett, A. Charnes, K. Paick		8. CONTRACT OR GRANT NUMBER(s) N00014-81-C-0236 N00014-82-K-0295
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Cybernetic Studies, UT Austin Austin, Texas 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 434) Washington, DC		12. REPORT DATE January 1984 13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Prior distribution assessment, Unimodality , Information theory, Maximum entropy

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we show how to take personally assessed information and use it to develop a continuous unimodal prior density function, perhaps for subsequent Bayesian analysis. The method is completely nonparametric and uses only the furnished information and no other. The technique is easily computerized and yields a closed analytical formula for the prior. The resulting distribution may be considered to be an inferential distribution.

END

FILMED

DTIC